RESEARCH CORES UNIT
INTRAMURAL RESEARCH PROTOCOL

A Conceptual Framework for Defining the Biological Domain of Applicability of AOPs

Author: Holly Mortensen, RCU Biologist HOLLY HOWELL Digitally signed by HOLLY HOWELL DN: c=US, o=U.S. Government, ou=USEPA, ou=Staff, cn=HOLLY HOWELL, dnQualifier=0000040393 Date: 2016.08.03 12:14:55 -04'00' 7/29/16
(Typed Name / Position; Signature) *Current Version Date*

A. Category of Research: 3 *(1, 2, or 3, as defined in Draft ORD PPM 13.7, Use of the Graded Approach)*

B. Is This Study Part of a Larger Project or Program Involving Other NHEERL Divisions, Other EPA Organizations, or Other Federal, State, or Local Agencies? Yes
If so, Who? NHEERL/ISTD, NHEERL/MED, NERL/HEADS

C. ORD Research Theme Addressed by This Project: CSS Systems
Project Number: 12.01 Adverse Outcome Pathway Discovery and Development
Task Number: 1.1c: Taxonomic Relevance of AOPs
Product Number: Product 1: A Conceptual Framework for Defining the Biological Domain of Applicability of AOPs

D. Approvals: RUSSELL OWEN Digitally signed by RUSSELL OWEN DN: c=US, o=U.S. Government, ou=USEPA, ou=Staff, cn=RUSSELL OWEN, dnQualifier=0000018156 Date: 2016.08.04 11:10:35 -04'00'

Director: _____
Russell D. Owen / Signature *Approval Date*

Branch Chief may assign additional reviews (e.g., peer reviewer, safety) either as a condition for final approval or to solicit advisory input (reviewer turnaround goal < 1 week):

Typed Name and Position	*Signature*	*Date*
Typed Name and Position	*Signature*	*Date*
Typed Name and Position	*Signature*	*Date*

Quality Assurance: THOMAS HUGHES Digitally signed by THOMAS HUGHES DN: c=US, o=U.S. Government, ou=USEPA, ou=Staff, cn=THOMAS HUGHES, dnQualifier=0000005997 Date: 2016.08.04 13:34:55 -04'00'
Thomas Hughes, RCU QA Manager *Date*

A. Table of Contents

B. Introduction

The purpose of this product is to define an organized framework for establishing the taxonomic domain of applicability for derived AOPs. This framework will be composed of a mainly computational set of tools, which will facilitate the assessment of similarities/differences of AOPs across species. Part of this effort is the testing and evaluation of computational tools, such as SeqAPASS, and others for the assessment of cross-species similarity at the molecular, biochemical, and pathway levels. This framework will establish the extent of conservation of various events defining the AOP across taxa (*e.g.*, MIE, KE) using quantitative measures of similarity for molecular events, but also address the establishment of rules for evaluation of taxonomic relevance for non-molecular events or events lacking more mechanistic information. Potential computational applications and related tools to be included in the framework will be evaluated through case studies of interest to program office partners, some of which are currently under review by the Organisation for Economic Co-operation and Development (OECD) Advisory Group on Molecular Screening and Toxicogenomics (EAGMST).

C. Purpose

The establishment of a taxonomic domain of applicability for a given AOP, with a focus on the mechanistic and molecular aspects, minimally requires the incorporation of publically available data on specific of MIEs, and KEs. The purpose of the work covered under this IRP is to guide research methods development and computational testing, as well as data integration and analysis of next generation data, to facilitate the understanding of the evolutionary origins of toxicological pathways. The objective is to provide a framework for work flow development and validation to maximize analysis reproducibility.

D. Test Substance:

Not applicable.

E. Methods and Experimental Approaches:

In order to understand the steps necessary in defining the interspecies domain of applicability for any given AOP, we have determined three AOP case studies based on the

presence/availability of information in the AOP wiki (Edwards 2014). With these three examples, we aim to cover the three situations that can occur relative to AOPs. These are: 1. *Broad Taxonomic Applicability of AOP*- the molecular initiating event (MIE) and associated cascade of key events (KEs) leading to an adverse outcome (AO) are highly conserved across species of concern; 2. *Limited Range of Taxonomic Applicability of AOP*- neither MIEs nor other pathway components are similar across species of concern; 3. *Lack of Taxonomic Concordance of AOP*- the MIE is conserved, but later KEs and the AO in the pathway are not similar across species of concern. The three case studies selected for analysis are AHR1 activation leading to developmental abnormalities (in birds), ER antagonism/Aromatase inhibition leading to reproductive disfunction (in fish), and Acetylcholinesterase inhibition leading to acute mortality (in fish). The following workflow description outlines the general experimental approach to characterize these case studies across taxa:

1. MIEs for each AOP are identified, as well as a small number of KEs, where each event is associated with a molecular target. This information is typically expert curated and obtained from the AOP wiki (Edwards 2014). When this is not the case this level of information will be obtained from the literature.
2. Characterization of the MIE across species in order to determine the most frequently studied organisms for each case study. An EPA in-house text mining technique (CSS, v-Liver) that implements MeSH (Medical Subject Heading) unique identifiers to quantify the number of citations of the MIE within the National Library of Medicine's PubMed citation system has been implemented.
3. Gene or protein sequence for each molecular target corresponding to an MIE or KE will be obtained for each species of interest. If information from multiple species must be used in the process of data analysis, the reference material and database (including access date) must be documented. Selection of highest quality and most appropriate data to address the research questions is important for obtaining relevant results.
4. The SeqAPass tool (Lalone et al. 2013) will be used to interrogate the molecular target for the most frequently identified species or group of species identified for each case study, in order to determine the level of variation (% sequence similarity) across MIEs and KEs between species.
5. Characterize the MIEs and KEs for each case study in terms of biological/ toxicological pathway networks. Selection of highest quality and most appropriate data to address the research questions is important for obtaining relevant results. Normalization steps may be needed to enable statistical analysis or data integration. Data processing, such as differential expression/abundance, or data integration steps if needed before final calculations should be applied to the all data unless there is a reason to apply processing on subgroups.
6. Determine network orthology (*i.e.* phylogenetic origin) to extrapolate shared adverse outcome between species of interest. The network topologies across species will be interrogated for the presence of known regulatory modules, especially those including toxicological pathways.
7. Make prediction (with level of confidence) of relevance for each case study. Evolutionarily conserved network topologies will be identified across species, and when appropriate, compared to H. sapiens or most sensitive species for relevance.

F. Equipment and Supplies

Hardware requirements:

>10GB of RAM
>250GB storage
Multi core processor (>= 4 physical cores)
Internet access is required

Software requirements:

R v3.2.1 or greater (RCoreTeam 2013)
Productivity software (for example, Office 2007)
Version control software (for example, Git)

G. Records: creation, inventory, and maintenance

All software and procedures for the data used will employ a version control system, such as Bitbucket (Outlined in Mortensen_GeneralDataAnalysisOP_5_18_16), or other methodology to capture computational steps in the analysis for transparency and reproducibility. Scripts used to manipulate the data leading to an output file for a subsequent step/ final product will include an output file that includes information regarding the date and description of key parameters used to create said file.

At the conclusion of a distinct research unit (such as work leading to a manuscript), a short report providing a summary of what was done (including the relevant file names of scripts and outputs) and findings will be completed. These reports shall serve as 'README' documents for the execution of the workflow in addition to the analysis of results.

H. Quality Control Requirements

Version and references for all libraries and software used will be documented in final publication. Verification of object format and output will be documented in script as appropriate. Use of example datasets provided by documentation for libraries or software, when they exist, will be used to verify functionality in the event that the method fails to perform as expected on work flow data. This data may also be used to aid in the identification of test dataset limitations.

This work relies on the use of existing, publicly available data. The data, source, and manner in which the data were obtained will be maintained in the records either as a script used to obtain the data or a README-type file housed in the directory containing the datasets. Any procedures which alter the data from its original form will be documented and a new file for the data created to ensure the original data is not altered. All data and variations will be stored in a shared drive to facilitate sharing. Computational files and scripts will be stored locally and files generated on the server or other high performance computer will be synced to the local folder for backup purposes. Copies of other files used in the workflow that are housed in other directories will be housed in the project folder to maintain organization.

Code review including at least one person who did not author the code will be performed prior to publication to ensure the reproducibility and readability of the code. Internal quality control of the analysis process relies on the use of external support (literature, expert knowledge) and verifying

identified relationships/results using those established in the literature.

I. Personnel assignments

Maureen Pittman, ORAU Student Services Contractor: methods development, data analysis and interpretation
Gerald Ankley, ORD/MED: consultation, interpretation
Carlie LaLone, ORD/MED: Task Co-Lead
Holly Mortensen Howell, ORD/NHEERL: Task Co-Lead, Product Lead

J. References

Edwards S (2014) AOP Wiki. In. http://aopwiki.org/

Lalone CA, Villeneuve DL, Burgoon LD, et al. (2013) Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action. Aquatic toxicology 144-145:141-54 doi:10.1016/j.aquatox.2013.09.004

RCoreTeam (2013) A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria.